Measuring quality of omnidirectional high dynamic range content

Anne-Flore Perrin[a], Cambodge Bist[b], Rémi Cozot[b], and Touradj Ebrahimi[a]

[a]Multimedia Signal Processing Group (MMSPG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[b]Institut de Recherche Technologique (IRT) b<>com, Cesson-Sévigné, France

ABSTRACT

Although HDR content processing, coding and quality assessment have been largely addressed in the last few years, little to no work has been concentrating on how to assess quality in HDR for 360° or omnidirectional content. This paper is an attempt to answer to various questions in this direction. As a minimum, a new data set for 360° HDR content is proposed and a new methodology is designed to assess subjective quality of HDR 360° content when it is displayed on SDR HMD after applying various tone mapping operators. The results are then analyzed and conclusions are drawn.

Keywords: omnidirection imaging, high dynamic range, tone mapping, subjective evaluation

1. INTRODUCTION

High Dynamic Range (HDR) is becoming a popular type of visual content in both professional and consumer markets. With advances in capture and display technologies, the HDR imaging pipeline is able to transmit potentially the full range of light information of a scene. This characteristic makes it overcome several physical and perceptual limitations of Standard Dynamic Range (SDR) imaging systems. Over the last three decades HDR content processing, coding and quality assessment have been a subject of attention in imaging.[1]

Omnidirectional or 360° imaging is an emerging format in the field of immersive multimedia. 360° cameras allow the capture of the entire field of view that covers a full sphere. This content is usually visualized using near-eye Head Mounted Displays (HMDs) allowing the viewer to freely change the direction of their sight across the omnidirectional scene. However, to the best of our knowledge, omnidirectional systems today are based on SDR imaging systems i.e. captured with a single exposure and viewed on a display with limited dynamic range. This makes current 360° imaging more vulnerable to the limitations of SDR systems especially from loss of information due to under and over exposures. This in turn could lead to a less immersive experience. HDR imaging seems to be a potential technology to resolve this issue. Therefore, in this paper, we set out to identify whether there is a qualitative benefit of applying existing HDR imaging workflows to the current SDR 360° imaging pipeline.

In order to investigate this issue, we generate our own omnidirectional HDR images with a commercially available 360° camera using classical techniques based on multiple exposure bracketing.[2] As the 360° HDR images have a higher dynamic range than that of typical head mounted displays, it requires additional processing to map the HDR image on to the display system. This step is known as tone mapping, and is implemented by a tone mapping operator (TMO). We consider well known off-the-shelf tone mapping operators for this work. The basis of this study is a qualitative experiment comparing the various TMOs with a single exposure 360° reference image. In order to conduct a rigorous evaluation, we propose a novel dual stimulus methodology designed specifically for pairwise comparison on an HMD. The conclusion of the experiment is that there is slightly improved perceptual quality of using a multiple exposure workflow for omnidirectional imaging. We identify this result as limitation of using consumer 360° camera. Based on our experiments, we also identify

Further author information: (Send correspondence to Anne-Flore Perrin.)
Anne-Flore Perrin: E-mail: anne-flore.perrin@epfl.ch, Telephone: + 41 21 693 4620
Cambodge Bist: E-mail: cambodge.bist@b-com.com, Telephone: + 33 25 635 8217

certain drawbacks with existing TMOs, and propose requirements for more robust and accurate tone mapping dedicated to 360° HDR. In addition to this, the novel dual stimulus approach of this study also gives new insights for estimation and extraction of regions of interest in omnidirectional images.

The remainder of the paper is outlined as follows. Section 2 presents the related work on HDR capture, tone mapping and subjective evaluation methodologies. The setup of the experiment is described in Section 3, and introduces equipment used, acquisition of a new dataset, content preparation and selection. The details of the evaluation methodology together with the experiment design are presented in Section4. Section 5 exposes in depth analysis of the results. Finally, we conclude the paper in Section 6.

2. RELATED WORK

In this section, the state-of-the-art literature related to this paper is reviewed. We begin with a discussion on HDR acquisition methods and related work done on omnidirectional imaging. This is followed by a brief review of tone mapping operators highlighting the lack of methods available for omnidirectional tone mapping. Lastly, we discuss previous work on subjective evaluation methodologies in both HDR and 360° imaging which serves as a platform for our proposed methodology.

2.1 HDR acquisition

High-end DSLR cameras (e.g., Nikon D5 and Canon 1DX Mark II) and specialized professional video cameras (e.g., Arri Alexa XT and Sony F65) are some of the state-of-the-art HDR capable capturing technologies available. However, most consumer cameras only capture 8-bit images and are far from capturing the full range of luminance. To overcome this limitation, multiple single photographs are taken at different exposures in order to capture details from the darkest to the brightest regions.[2] In this study, we consider a traditional single exposure workflows based on multiple exposures as seen in Figure 1.

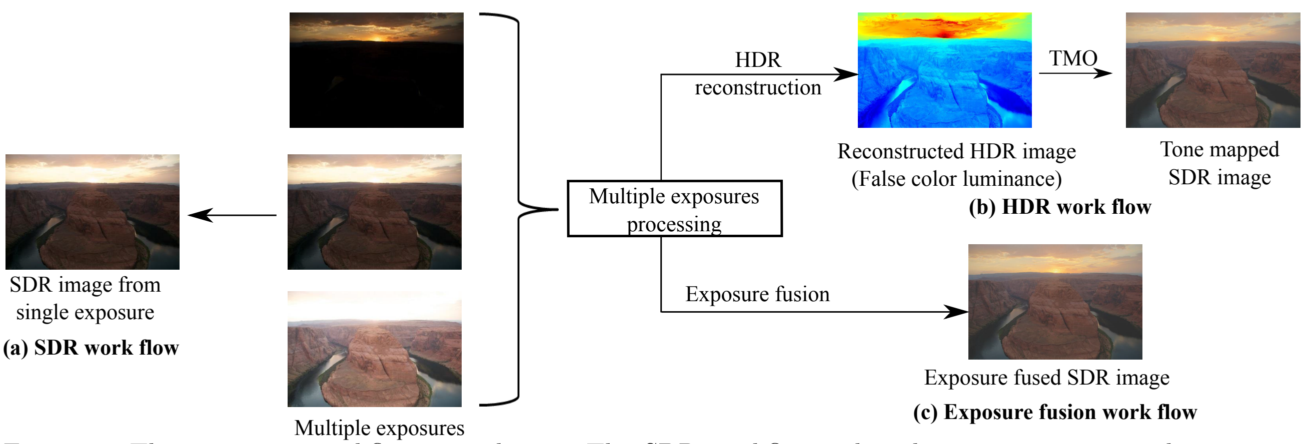


Figure 1: Three imaging workflows are shown. The SDR workflow is based on capturing a single exposure image. The example in (a) shows the process by capturing the mid-exposure of the scene. The HDR workflow in (b) describes the typical HDR pipeline consisting of bracketing, HDR reconstruction and tone mapping. The reconstruction technique used in this paper is based on the work of Debevec and Malik.[3] This is one of the first methods that proposed to combine multiple exposures to produce an HDR image with pixel values proportional to the physical radiance values of the light captured. After reconstruction we employ the tone mapping step. The Exposure fusion workflow as seen in (c) is based on the 'HDR mode' we find in mobile devices. This process is akin to the work by Mertens et al.[4] which consists of blending multiple exposures into a single image by skipping the HDR reconstruction step.

Although omnidirectional HDR imaging is not yet available on consumer multimedia platforms, it is very much present in Image Based Lighting (IBL) applications. IBL is a 3D rendering method used in computer graphics for illuminating 3D objects with real world illumination.[5] This often requires capturing an omnidirectional image (typically using multiple exposures for HDR), mapping the HDR illumination on to an environment map and

Table 1: Description of various operators used in our experiments.

Operator	Type	Description
Linear TMO	Global	Simple linear scaling involving the normalization of the HDR radiance values between [0 1] followed by the application of a gamma correction of 2.2.
Photographic TMO[9]	Global	Based on the work by Reinhard et al.,[9] this TMO relies on photographic principles. In our experiments, we applied only the global component of this algorithm mainly to compress high luminance values.
Display adaptive TMO[10]	Global	Display adaptive tone mapping preserves the contrasts of the HDR image taking into account the characteristics of an output display. We set the output display specification similar to that of an HMD, i.e. up to 100 nits in peak brightness and surround environment of 0 lux.
Detail preserving TMO[11]	Local	This TMO is based on a contrast domain processing algorithm designed to preserve details in the HDR image.
Exposure fusion[4]	Local	This technique fuses multiple exposures into a single image giving similar results to that of a TMO. However, as it never generates an HDR image it is not considered an HDR technique.

placing 3D objects in this environment to recreate real world lighting conditions for the objects. Typical IBL applications are in the film and the video game industry.[6] Capturing content for IBL is often a cumbersome and expensive process which requires professional DSLR cameras, motorized rigs, stitching software and dedicated processing.[7] To avoid this, we work with a consumer omnidirectional camera from which we take multiple exposure images to create our own 360° HDR images. Furthermore, this choice also positions this study to tackle a real world scenario using consumer 360° cameras and consumer HMDs.

2.2 Tone mapping operators

The development of tone mapping operators has been an active field of research for many years. The goal of a TMO is to reduce the HDR contents dynamic range to match that of an SDR display. Simple TMOs are based on operations such as scaling, clipping or gamma correction. More complex TMOs take into account the characteristics of the scene or properties of the human visual system in order to reproduce the best visual experience on an SDR display. TMOs are broadly classified into global and local operators. Global operators apply a single mapping function to every pixel in the image while local operators apply unique mapping per pixel in an image depending on its neighboring pixels. For this study, we have chosen commonly used TMOs that have been freely available on commercial software[8] for many years and are fundamental to HDR imaging. A description of various TMOs along with the exposure fusion method is described in Table 1.

The TMOs considered in this study are based on 2D imaging. This is mostly because very little has been done in omnidirectional HDR tone mapping. Amongst the few works in literature, Hausner and Stamminger[12] present and extension of the Photographic TMO which adapts to the user's central field of view by using tracking information. Yang et al.[13] and Mikamo et al.[14] display two different tone mapped images of the same HDR input image on each eye of a binocular display (such as an HMD). When seen through a binocular display, the fused image presents more visual richness and detail than both tone mapped versions. These works are potential directions for research in omnidirectional HDR imaging for HMDs but they are out of the scope of this study. Hence, we will only consider existing TMOs part of existing 2D HDR workflows for qualitative assessment against SDR 360° workflows.

2.3 Quality evaluation for HDR and omnidirectional imaging

Several studies in the past have evaluated the visual quality of tone mapping operators. Ledda et al.[15] conducted a subjective evaluation of tone mapping operators with a reference HDR display while Yoshida et al.[16] presented a study evaluating tone mapped images with real-world scenes. Kuang et al.[17] did a pairwise comparison experiment between TMOs on a single SDR display. Evaluations of tone mapping for HDR videos have also been well covered by Eilertsen et al.[18] and more recently by Melo et al.[19] Despite the vast study done on this topic, the underlying conclusion amongst existing works is that preference of TMOs is very subjective. In a subjective evaluation more aligned to our work, Narwaria et al.[20] compared tone mapped images with single exposure images. The study concluded that observers saw no significant differences between tone mapped and single exposure content. The authors explain this unexpected result citing a number of perceptual cues including details in the bright or dark areas, unnatural colors, overall contrast, naturalness of the scene, etc. We will also consider these perceptual factors for our experiments.

Unlike HDR imaging, limited work exists on qualitative assessment of omnidirectional imaging. Yu et al.[21] proposed two objective metrics for omnidirectional videos comparing different panoramic projections. Zakharchenko et al.[22] proposed a quality metric which remaps omnidirectional images to a craster parabolic projection to compare different geometrical projections. Upenik et al.[23] introduced a testbed for single stimulus subjective evaluations of 360° images on HMDs. We build upon this testbed to introduce new dual stimulus evaluation method for omnidirectional imaging in order to conduct our experiments.

3. EXPERIMENT SETUP

This section introduces the experimental setup by describing the equipment used, the generation of a new HDR 360° dataset, the content preparation and selection.

3.1 Equipment

The content for our experiments was acquired using the Ricoh S 360° panoramic pocket camera. This consumer camera provides 8-bit, 5376x2688 (14MP), RGB pictures. The good quality stitching when generating a single 360-by-360 degree sphere image is due to its two 190-degree field of view cameras. The Ricoh theta S allows the use of a shutter speed ranging from 1/6400 to 1/8 seconds and an ISO sensitivity ranging from 100 to 1600 with a wide lens aperture (F2.0) and a focal length of 1.31. An automatic exposure control is able to acquire images within a range of Exposure Values (EVs) from -2 to 2 by steps of 1/3. Each content is stored as an equirectangular projection compressed version. The used compression is a DCF2.0 and Exif ver. 2.3 compliant JPEG process, with a quality factor of 95. To ensure the stability of the camera while acquiring the multi-exposure pictures, a tripod was used, when not exploiting the environment itself to hold the camera.

The rendering and evaluation of the content produced are based on the subjective evaluation testbed implemented for omnidirectional images proposed by Upenik et al.[23] This testbed enables test sessions proceeding as follows: textual instructions, a training session and the subjective evaluation using a single-stimulus methodology consisting of displaying a stimulus, then a scoring menu after interaction via a one touch button and finally the selection of the score before continuing to the next stimulus. The testbed acquires information from every stimulus including the set of each subject's scores, the tracks of the directions of view and the timestamps of the stimulus visualization. The direction of view is recorded by three coordinates, yaw, pitch and roll, representing the angles formed with the normal, lateral and longitudinal axes, respectively. The frequency of acquisition is about 60 Hz and the precision of timestamps is 10^{-7} seconds. The equipment required to perform the evaluation is a hand-held device (e.g. iOS mobile) combined with an HMD.

We have used the iPhones 6 and 6S, compatible with the above described testbed. The screens in those devices are 4.7 inch large in diagonal with an HD resolution (326 ppi). Their maximum brightness is 500 nits and their contrast ratio is 1400:1. The color space representation of displays are full sRGB. There is no difference between the two devices displays and the calibration process followed was strictly the same. To the best of our knowledge, there is no peak brightness recommendations or standards for HMDs. Therefore, to insure the subjects comfort as well as an optimal visualization setup, it was decided to set the display luminance to 100 nits. This decision is based on a few pilot tests as well as on recommended brightness settings for SDR TV in broadcast.[24] The peak brightness was determined using a white screen and by manually adjusting the iphone's brightness slider. Measurements were taken using the x-rite i1 Display Pro[*] and the i1 Profiler Software 1.1.1.

[*] http://www.xrite.com/categories/calibration-profiling/i1display-pro

(a) EV 2 (b) EV 1 (c) EV 0 (d) EV -1 (e) EV -2

Figure 2: Multi-exposure picture of the *Lake* content

(a) Linear TMO (b) Photographic TMO (c) Display adaptive TMO (d) Detail preserving TMO (e) Exposure fusion

Figure 3: Tone-mapped and exposure fusion pictures of the *Lake* content

The experiments were conducted using the Merge virtual reality[†] headset. This HMD is compatible with Android and iOS smartphones. Its simple double input interfaces (referred to as buttons in the following) facilitate and expand possible interactions within the testbed. The adjustable lenses lead to a more comfortable experience as they are designed to fit one's specific eye distance. This HMD guarantees a field of view of 90 degrees. The entire system has a precision of 8.3 pixels per degree.

3.2 HDR omnidirectional dataset

The 360° HDR dataset is composed of 43 contents. Each content representation includes a set of 5 multi-exposures pictures, an HDR reconstruction as well as tone-mapped and exposures fusion versions of the scene. The multi-exposure pictures were captured using the consumer camera Ricoh Theta S, described in section 3.1, with EVs ranging from -2 to 2 by step of 1. As an example, Figure 2 presents the five multi-exposure images of the scene *Lake*. A large part of the wide range of luminance of the scene is thus captured. More precisely, lowest EVs capture the details in bright areas while highest EVs acquire the details in dark areas. Figure 3 illustrates the SDR images resulting from the four TMOs and Exposure fusion described in section 2.1.

The captured contents are classified in three content types, namely *Indoor* (9 contents), *Outdoor* (24 contents) and *Night* (10 contents). This dataset aims at the evaluation of a broad collection of scenes in addition to capturing a wide variety of shooting conditions and environments e.g. with various overall contrast and brightness, diffuse and specular reflections, deep shadows and bright highlights. A specific attention was devoted to have as few as possible multi-exposure bracketing impairments such as ghosting effects and noise. The introduced dataset is publicly available[‡]. It should be noted that even though 5 exposures enable the capture of a large part of the dynamic range of a scene, the sensor properties of the Ricoh Theta S consumer camera are limited. We shall revisit this issue in Section 3.4.

3.3 Content preparation

After capturing different exposures, there are several ways to generate an HDR image. Based on the study of Akyuz et al.,[25] state-of-the-art reconstruction algorithms are not statistically superior to each other in terms of accuracy. Hence, we simply apply the classical method proposed by Debevec and Malik[3] using a triangular weighting function for our study. For all the TMOs, we applied the recommended settings suggested by the authors except for the Display Adaptive TMO where we applied the surround lightings setting specific to our HMD. It must be noted that all pre-processing were done on the equirectangular projection of the contents.

A special attention should be paid to processing equirectangular projections with *Detail preserving TMO* and *Exposure fusion*. This is because both operators implement a local processing which implies that the left and right sides of the projected image will have a different processing and as a consequence a vertical line can appear on the HMD, clearly depicting the region where the stitching is applied. This is a major drawback of working

[†] https://mergevr.com/goggles

[‡] http://mmspg.epfl.ch/360hdr-consumercamera

in the equirectangular projection and creates a reoccurring artefact for all local processing potentially affecting users judgment during subjective quality assessments. We resolved this issue by doing additional processing for local operators. This involves three steps: 1) concatenating the columns of left part of the projection to the right part and vice versa, 2) applying the local operator and 3) removing the additional columns and retrieve the locally processed image.

3.4 Content selection

In order to conduct a reliable subjective quality assessment, the test images should contain a large variety of stimuli. In HDR imaging, scenes with extreme variations in lighting levels are often considered. Keeping this in mind, we have selected 8 different HDR images (shown in Figure 4 along with log2 histograms) which are representative of a diverse set of situations such as indoor, outdoor and night scenes with varying lighting conditions. Considerations about having minimal artifacts (e.g. ghosting, noise and over/under exposure) were also taken in account. The Table 2 summarizes the content selected and presents their characteristics.

Table 2: Characteristics and global statistics over the entire image

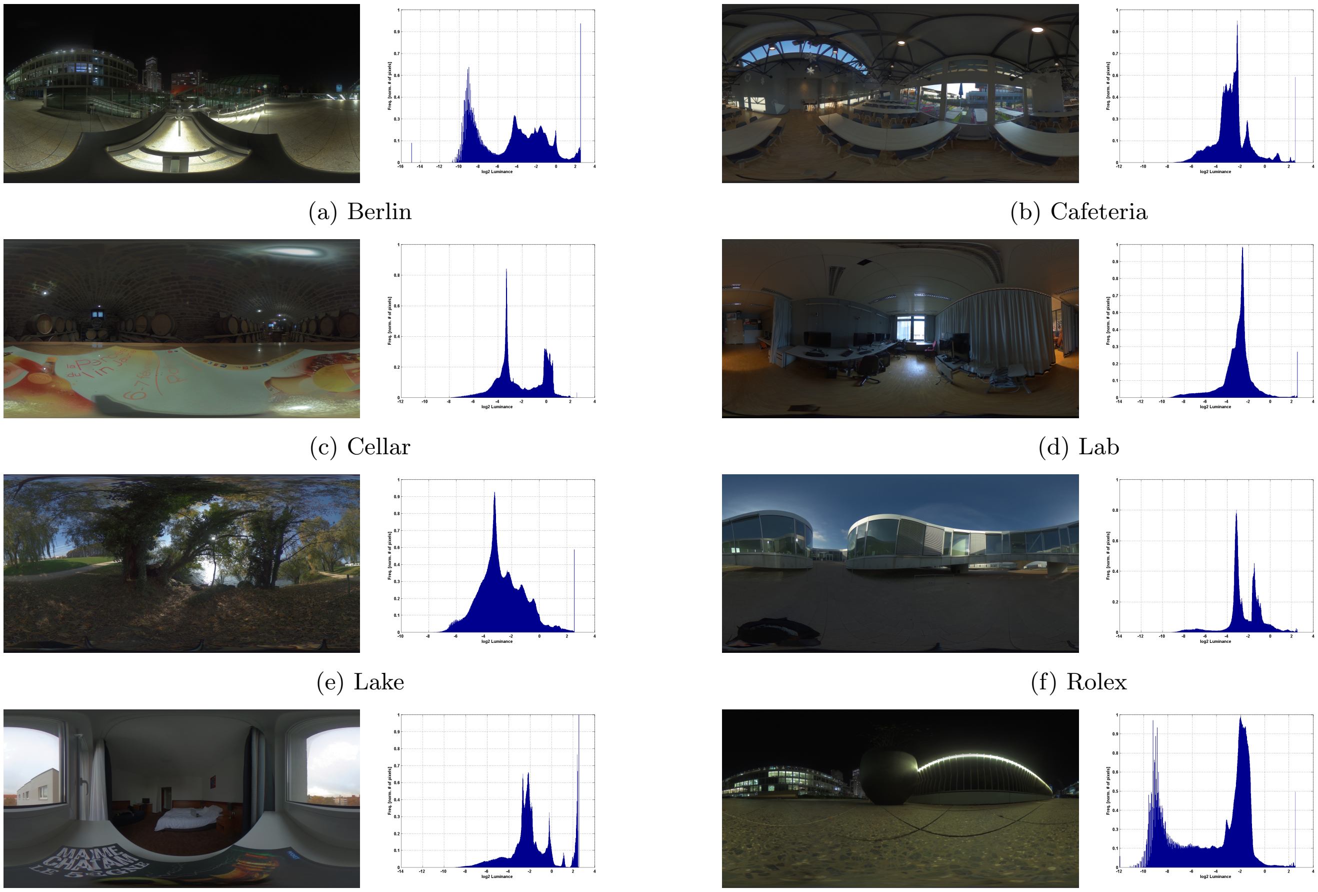
		Reference in database		Exposure time		Statistics		
Set	Name	Content type	Reference	min	max	DR	Key value	SI
Training Set	Trail	Outdoor	01	1/6400	1/400	7.92	0.46	0.21
	Bridge	Outdoor	05	1/6000	1/125	7.59	0.46	0.12
	Sculpture	Outdoor	21	1/4000	1/200	7.93	0.59	0.15
Test Set	Room	Indoor	02	1/200	1/30	9.88	0.54	0.10
	Lab	Indoor	04	1/30	1/8	9.70	0.54	0.14
	Cellar	Indoor	05	1/180	1/30	7.86	0.55	0.10
	Cafeteria	Indoor	09	1/1500	1/40	8.22	0.41	0.11
	Vase	Night	01	1/30	1/10	10.59	0.52	0.19
	Berlin	Night	02	1/60	1/8	12.27	0.42	0.26
	Lake	Outdoor	02	1/5000	1/160	7.97	0.44	0.23
	Rolex	Outdoor	23	1/6400	1/500	9.58	0.57	0.13

We considered a number of different global statistics over the entire sphere for each image. The statistics for the experimental images can be found in Table 2 and are defined below:

1. Dynamic Range (DR):
$$DR = log_2(L_{max}/L_{min})\,, \tag{1}$$
where L_{max} and L_{min} are the highest and lowest scene referred luminance values clipped at 0.1 and 99.9th percentile to make the metric robust against outliers. This is the most popular unit for expressing dynamic range and is measured in f-stops.

2. Key Value:
The key value is a popular metric in HDR imaging to indicate the average brightness of a scene[9] as depicted in the equation below:
$$Key\ value = \frac{logL_{avg} - logL_{min}}{logL_{max} - logL_{min}}\,, \tag{2}$$
where $logL_{avg}$ is the log geometric mean of luminance of the image, L_{max} and L_{min} are the maximum and minimum scene referred luminance values clipped at 0.1 and 99.9th percentile.

3. Spatial Information (SI):
The SI is an indicator of spatial details in an image. This was determined by first applying the Sobel filter on the linear TMO image followed by calculating the output image's standard deviation.



(a) Berlin (b) Cafeteria (c) Cellar (d) Lab (e) Lake (f) Rolex (g) Room (h) Vase

Figure 4: The equirectangular projections and the histograms of the images used in the experiments. The linear TMO has been applied for each of the images above. The histograms are computed over the entire sphere of the omnidirectional image, showing the relative frequency of the log 2 luminance of the pixels.

In Table 2, we see that the various statistics don't show a clear indication of the large variety between selected images. One of the main concerns of the dataset is that the dynamic range of the images varies from 7.59 to 12.27 f-stops. We believe this limitation to be due to the Ricoh Theta S sensor at extreme EVs. In fact, one can observe over-exposed and under-exposed regions in the reconstructed linear HDR scenes which is atypical in HDR imaging. Nevertheless, this dataset remains useful as it is the best possible dynamic range that can be captured using the current consumer omnidirectional cameras which is the use-case we are addressing.

This is not to say that content is not diverse. To put into perspective the diversity of the selected omnidirectional content, we refer to the work in video tone mapping by Boitard et al.[26] If we consider the view-port as video frame and the movement of the observer as the motion of the camera, it is possible to study the changed key value as the user navigates across the 360° image. To simulate this, we extract the view-ports along the center of the omnidirectional images and calculate the key value per view-port as seen in Figure 5. It must be noted that the brightness not only varies between images but also changes significantly within the entire sphere in each image.

4. EVALUATION METHODOLOGY

In this section is discussed the design of the subjective test, from the design of a methodology for omnidirectional content pair-comparison to the definition of our test and the creation of a comprehensive questionnaire.

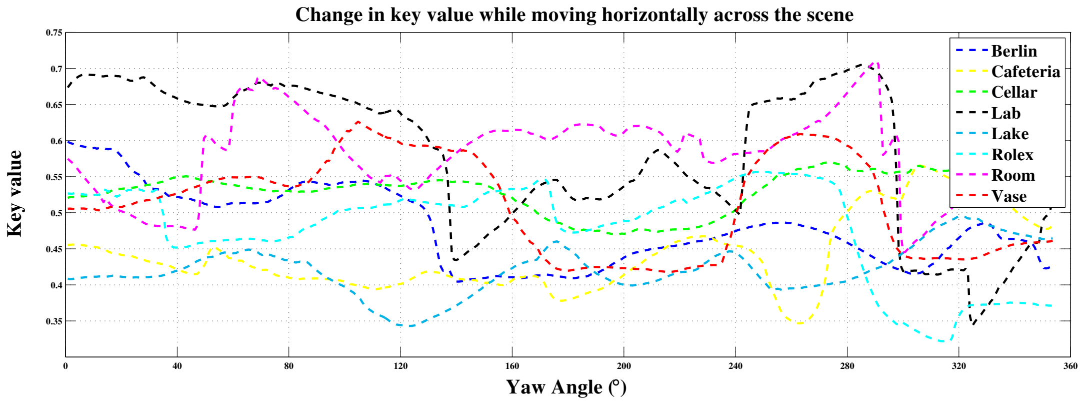


Figure 5: The variation in key value of the omnidirectional HDR content used in experiments. The key value is calculated using Equation 2 for a given view-port across the entire yaw angle of the scene while fixing the pitch and roll to 0°. This graph shows the content selected was diverse with varying luminance levels to challenge TMOs.

4.1 Pair comparison

To the best of our knowledge, only the single stimulus(SS) methodology has been used in previous subjective tests for omnidirectional content. A single stimulus evaluation consists in the presentation of a single content which is then assessed according to a predetermined scale. In pair comparison (PC), the relation between two images or image sequences is evaluated. The set of stimuli is usually presented juxtaposed or sequentially. Even though those are two preferential assessment methods, as emphasized by the ITU-R recommendation,[27] the choice of using SS or PC is based on the context of the analysis and the aim of the experiment. SS sorts the assessed impairments in an absolute way while PC permits to discriminate one when compared to another, by determining their relative quality. As we conduct here an analysis on the relative improvement of the perceptual quality of multi-exposures over single exposure workflows, a PC methodology is more suitable.

A number of methods of pair comparison methodologies were envisioned. This included well known side-by-side implementations such as butterfly and split screen. The main disadvantage of these approaches are lack of naturalness of the scene especially when seen through an HMD as well as the introduction of large amplitude head movements potentially leading to sickness effects. In Figure 6, we demonstrate the limitations of various side-by-side implementations for the *Room* image. Hence, we propose an alternate PC approach based on image toggling.

4.2 Toggling

A pair comparison toggling approach allows users to visualize both contents by switching between them. We build on the testbed described in section 3.1 to introduce a new double stimulus assessment methodology for

(a) Split Screen (b) Horizontal Butterfly (c) Vertical Butterfly

Figure 6: Approaches considered to reproduce side-by-side pair comparison methodology in an omnidirectional environment are presented. Using a split screen as seen in (a), forces the user to evaluate different parts of the same content which violates the construct validity of the experiment. Similarly the butterfly comparison in (b) and (c) will results in very unnatural environments which may bias the assessment. Additionally, comparing contents in (c) will require large amplitude head movements, likely to generate sickness effects. Thus, the toggling approach for the pair comparison evaluation is proposed.

omnidirectional content. The proposed pair comparison evaluation starts by displaying the test stimulus first. By pressing the right button of the HMD, one can display the reference stimulus for the same view-port. By pressing the right button multiple times, the user can toggle between the test and the reference stimuli. Each scene is labeled either as T for test or R for reference at the center of the view-port, indicating to the subject which stimulus is currently displayed. In our experiments, it was mandatory to toggle between the reference and the test stimuli at least twice before scoring. No maximum limit of the number of toggling was set. The vote is enabled only on the test stimulus so that the last visualized content is the one to be assessed. The left button of the HMD is used for voting. By first pressing the left button a vote menu is displayed. With the help of a red cursor at the center of the view-port, users can select their score. Once the cursor has been correctly positioned over the preferred score, the user can cast his/her vote by pressing again on the left button. The order of the presentation of pair comparison stimuli are randomized in such a way that never the same content is assessed successively.

The proposed double stimulus testbed also collects similar data to single stimulus testbed such as scores and view direction tracking. It also records the toggling information by storing the timestamp and the view-port direction when users toggle. The analysis of such information can be used to investigate comparison process and to observe the regions of interest for HDR omnidirectional contents. This could potentially help identify subjects behavior during an assessment, gather data for design and implementation of future omnidirectional TMOs, conceive and validate new subjective quality assessment methodologies and objective metrics.

4.3 Experiment design

An adjectival categorical rating methodology was selected on a 5-point grading scale to score pair comparison results. The assessment scale stands as follows: *1: T worse than R* ; *2: T slightly worse than R* ; *3: T same as R* ; *4: T slightly better than R* ; *5: T better than R*. T refers to the test stimulus (TMO or exposure fusion image) while R is the reference stimulus. The rating scale is displayed on the voting menu of the testbed. The reference image is chosen as the single exposure image with EV = 0. This choice is based on photographic principles as the mid-exposure usually permits the acquisition of bright and dark areas without favoring neither highlights nor shadows.

Additionally, we created a post-questionnaire investigating the following criteria:

1. Criteria considered during the evaluation:
   HDR images are evaluated according to various criteria, such as image contrast, naturalness, colorfulness and overall brightness. In view of recommendations for the development of omnidirectional TMOs as well as for future questionnaires appraising 360° HDR contents, an investigation of the criteria impacting the assessment is conducted. Narwaria et al.[28] reviewed the criteria on which is based the differentiation of tone mapped images. The recurrent and most used criteria were selected for our questionnaire and are the following: *Details in the bright areas*, *Details in the dark areas*, *Unnatural colors*, *Ghosting*, *Noise*, *Overall brightness*, *Overall contrast* and *Naturalness of the scene*. These criteria were ranked from the most considered (1) to the least considered (8) when making the evaluation decision.

2. Interest and appreciation of the content:
   The assessment of the interest and appreciation of the content is a self-made questionnaire investigating if a content is *Boring*, *Interesting*, *Colorful*, *Aesthetic*, *Familiar*, *Of quality*, *Pleasurable* and *Immersive*. The subjects select the attributes when they find it suitable. The selection of several adjectives is allowed. Those dimensions are selected based on the information we want to gather to explain the possible discrepancies in the results we foresaw and are inspired from the work of Mansilla[29] on Quality of Aesthetic Experience.

3. Sickness:
   In order to accurately evaluate the degree and type of sickness generated by our experiment, the widely used simulator sickness questionnaire developed by Kennedy et al.[30] is included in our questionnaire. The symptoms *General discomfort*, *Fatigue*, *Headache*, *Eye strain*, *Difficulty focusing*, *Increased salivation*, *Sweating*, *Nausea*, *Difficulty concentrating*, *"Fullness of the head"*, *Blurred vision*, *Dizziness with eyes open*, *Dizziness with eyes closed*, *Vertigo*, *Stomach awareness* and *Burping* are evaluated on a 4-point grading scale from none (1) to severe (4).

4. Virtual reality experience:
The aim of the such questions is to acquire sufficient knowledge on the experience to improve the next design of the evaluation methodology. An investigation of the overall appreciation of subject's experience in terms of *Immersion*, *Isolation*, *Enjoyment/Valence*, *Arousal/Excitation* and the *Enjoyment due to the novelty of the visualization* is conducted. A 5-point scale from strongly disagree to strongly agree is used when asking subjects if they experienced the above-mentioned attributes. The immersion and isolation attributes evaluate the interest in using an HMD for the evaluation, the valence and arousal attributes come from the Self-Assessment Manikin[31] measuring the emotions generated by the experiment. Finally, the novelty effect could have an influence on our results. It could be interesting to know if subjects are aware of this bias during the assessment. Also, the level of experience of subjects regarding their use of HMD is gathered. An additional open question also asks them to describe their VR experience in a few words.

The study is carried out on eight contents, selected from the dataset introduced in section 3.2 following the process described in subsection 3.4, and on four TMOs and an exposure fusion operator, described in section 2.2. Overall, 40 pair comparison stimuli are evaluated. A single test session of about 20 minutes is needed for the assessment of all pair comparison stimuli.

Prior to the test, subjects were screened for correct color vision and vision accuracy using Ishiara and Snellen charts. As recommended in the simulator sickness questionnaire,[30] subjects were asked if they are in a normal state of health. Any violation of the previous conditions results in the rejection of the subject from the experiment. Overall, 25 (17 male and 8 female) subjects participated in tests and were 29.5 years old in average, ranging from 19 to 55 years old. After being instructed with the study process, subjects were requested to read the information and consent forms and the post-questionnaire, as well as an explanation sheet, describing complex terminologies used in the questionnaire. Any question was answered to before starting the experiment.

In order to have subjects familiarized with the assessment procedure as well as to reduce across-subjects grading discrepancies, instructions are again provided by the testbed and a training session is performed in advance of the test session. Three pair comparison stimuli were rendered during the training session in the following order: The photographic tone mapped *Bridge* content, assessed as worse than R (1), the display adaptive tone mapped *Trail*, evaluated as equivalent to R (3), and the exposure fusion of *Sculptures*, assessed as better than R (5). Different contents were used in the training session when compared to test session in order to prevent the introduction of any bias. The test session starts right after the completion of the training session. As mentioned previously, subjects had to toggle at least twice for each pair comparison and could only vote on the test stimulus T. No time-constraints were defined per stimulus or per test session. Once the test session is over, subjects are required to fill the post-questionnaire. The test was carried out in a calm environment, free from any disturbances. A rotatable chair is used during the assessment for subjects' comfort as well as to ease their 360° navigation within the omnidirectional content.

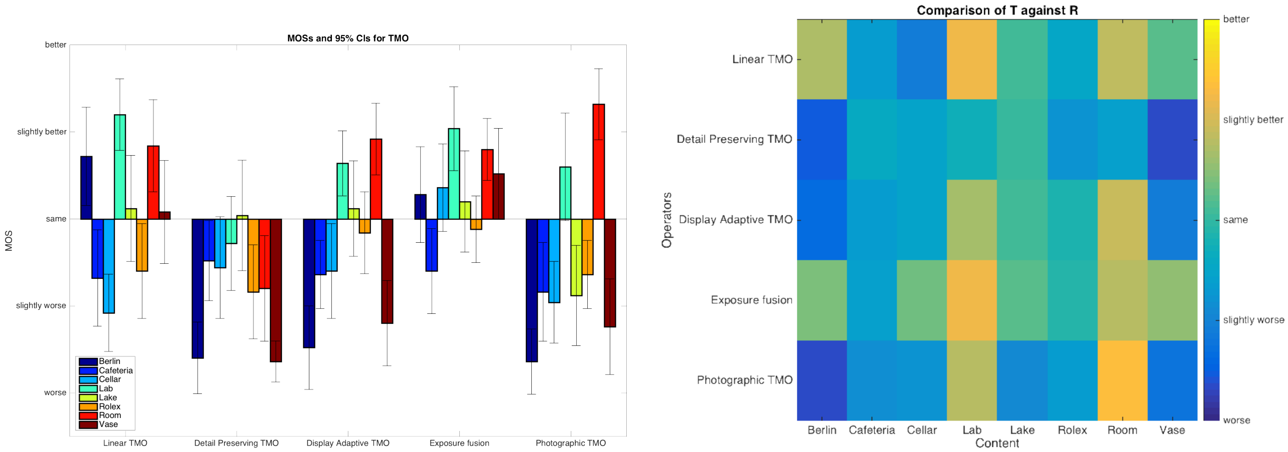
5. RESULTS AND ANALYSES

The analyses cover three types of information, namely, the subjective scores, the post-questionnaire data and the tracking of the direction of view combined with toggling information.

5.1 Subjective scores

After an outliers detection based on ITU-R recommendations BT.500[27] , which did not reject any subjects of this study, Mean Opinion Scores (MOSs) and 95% Confidence Intervals (CIs) were computed as depicted in Figure 7. Figure 7a shows the variations across contents, for each operator considered while Figure 7b presents a comparison matrix depending on contents and operators.

Overall, the results do not show a clear preference of the processed content over a single exposure content. Indeed, MOSs mainly range from 2 (T slightly worse than R) up to slightly above 4 (T slightly better than R), indicating a moderate improvement or deterioration of the perceived quality. Based on the Figure 7b, the exposure fusion and the Linear TMO show better performance when compared to the other operators. The extent of the performance is a similar or slightly improved perceived quality. On the other hand, the Photographic,



(a) MOSs and 95% CIs per content per operator (b) Comparison of MOSs

Figure 7: MOSs and CIs analysis

Display Adaptive and Detail Preserving TMOs do not perform as well. This result does not come as a surprise and can be explained through the limitations of the consumer camera. Let us recall that the DR measure of the dataset spanned from 7.59 - 12.27 f-stops. These TMOs are designed for content with a DR greater than 16 f-stops. Thus, this result suggests that TMOs require further evaluation using professional 360° HDR content having higher DR. The dataset is ideal for the exposure fusion algorithm which works well for images consisting of equally spaced EVs. As this method avoids the HDR reconstruction step and is thus independent of the DR of the scene. Furthermore, it is well known in HDR imaging that a linear TMO often results in a dark image and loss of detail.[32] The fact that we can clearly visualize the Figure 2a and observe little loss of detail also indicates the lack of DR in the content.

In addition to this, we can observe significant differences across contents, based on several non-overlapping CIs. This illustrates the variety of the chosen contents. It is worth noting that the variation of key value across omnidirectional content is a relevant indication for content selection. All operators, except the detail preserving TMO, are particularly preferred over the reference for the contents *Lab* and *Room*, two indoor contents with an outside view through a window. No trend in the statistics was found in Table 2 and Figure 5 seems to justify this behavior. This demonstrates the need for a precise questionnaire investigating specific criteria.

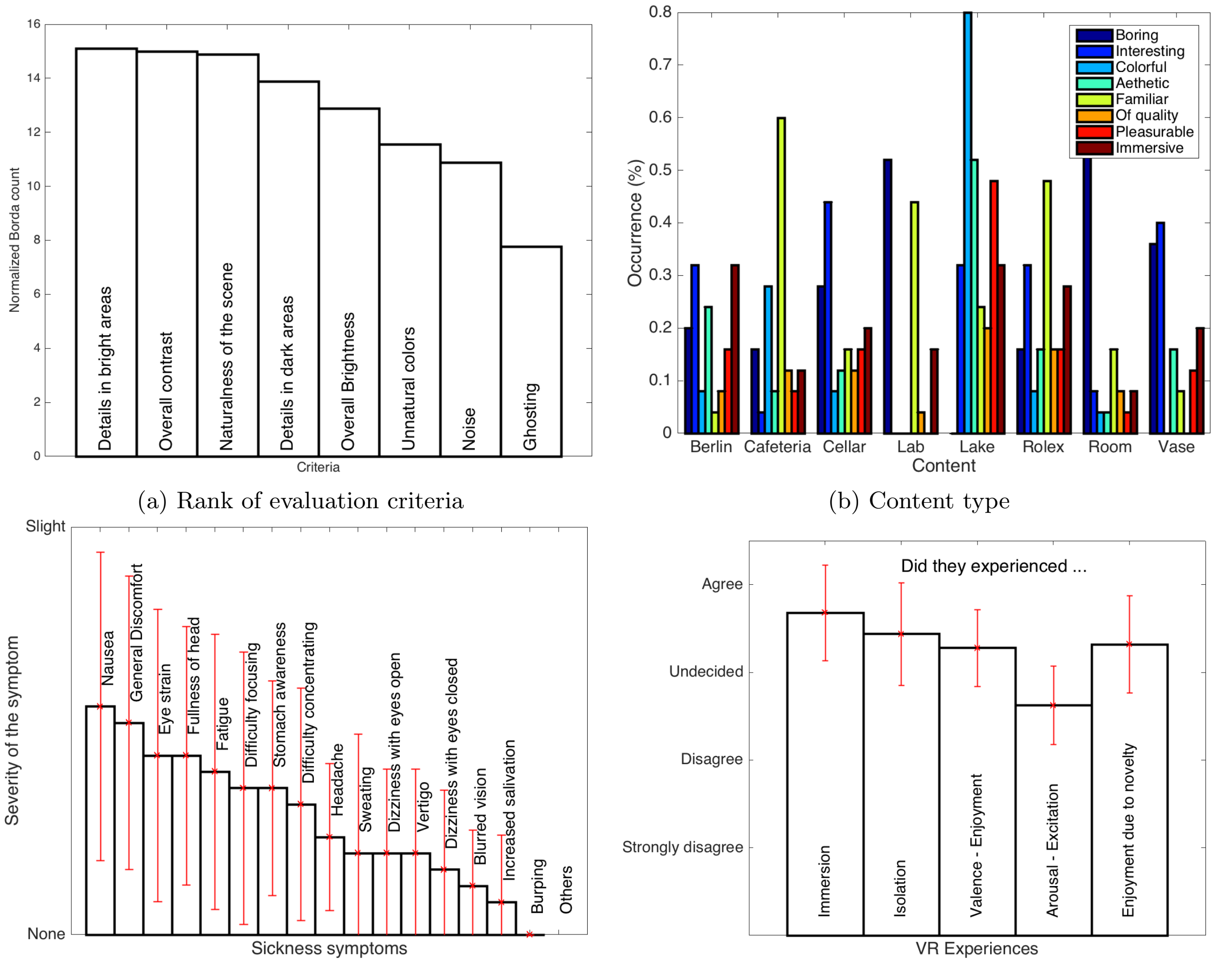
It should be noted that the two night contents have similar noise properties as they both have been shot in low light conditions. This affects the score chosen by the subjects on various TMOs. Most operators were assessed as slightly worse or worse for the night contents while the Linear TMO and exposure fusion scored between same and slightly better. This result emphasizes the importance of noise-aware tone mapping needed in omnidirectional HDR imaging just as much in 2D HDR imaging.[33]

The non-normality of scores distributions prevents us to run a repeated measure ANOVA, especially considering our sample size (=25), not sufficient to overcome violation of assumptions. As the aim of this paper is not to differentiate TMOs but to investigate the need of a dedicated TMO for omnidirectional contents, no non-parametric test have been performed as they would be hard to interpret in the context of this study.

5.2 Post-questionnaire answers

5.2.1 Criteria considered during the evaluation

In order to identify which criteria have the most impact on the assessment, the Borda count method[34] was applied. The results are presented in Figure 8a. From the latter we see that the details in bright areas, overall contrast and naturalness of the scene are the most important factors for the subjects. Those criteria should be considered in a questionnaire for future subjective quality assessments.



(a) Rank of evaluation criteria

(b) Content type

(c) Sickness symptoms

(d) VR experience

Figure 8: Post-questionnaire analysis

5.2.2 Interest and appreciation of the content

The Figure 8b illustrates the subjects choices of suitable adjectives describing the contents. It has to be noted that the contents showing the best results, *Lab* and *Room* were both assessed as boring by more than 50% of the subjects. This is an interesting finding that could be further investigated regarding subjective tests content selection. When considering the content *Lake*, its MOSs results indicate a similar perceived quality of T and R stimuli for all TMOs except for the Photographic TMO. The content was assessed by more than 80% of subjects as colorful and the above-mentioned operator introduced unnatural colors. Even if the unnaturalness of colors criteria was among the least considered, we can conclude that this criterion is still of importance when tone mapping omnidirectional HDR content. This fact emphasizes the importance of the criterion *unnatural colors*, though it is assessed as the 6th ranked criteria in terms of importance by subjects. We thus suggest to include all the investigated criteria in future test questionnaires.

5.2.3 Sickness

Various works in literature have reported that a virtual environment can generate sickness[35] . A subjective evaluation is likely to be biased if users are in discomfort or tired. Sickness effects are included in subject's discomfort. The questionnaire aims to certify the validity of the test. The Figure 8c reports the degree of severity of the sickness effect of the subjects. We can observe that the extent of sickness is from none to slight. In addition, 48% of our subjects reported not having experienced sickness. These two facts confirm the proper design of the test for pair comparison, concerning the sickness effect. Overall, when considering the severity and occurrence of sickness symptoms, Nausea and General discomfort characterize the sickness experienced by our subjects.

5.2.4 Virtual Reality Experience

The investigations carried out on subjects' experience and feedback are reported here. Prior to the test, 36% of subjects had never used an HMD while 16% had used it more than 5 times. We can conclude that our population sample is representative of the population when considering their VR experience.

The evaluations of subjects experience show a moderate immersion, isolation and enjoyment. However, the subjects acknowledged a lack of arousal. Those results are mainly explained in the open question which asks for summarizing subjects experience, at the end of the post-questionnaire. The justification of the above comment was provided by more than a quarter of the subjects mentioning that their excitation level was counter-balanced by their perception of the low level of quality of the images, especially when pixelation was perceived, or by the experience of sickness.
The reported feedback from the subjects in the open question also indicate discrepancies in the appreciation of enjoyment and perceived quality. Spontaneously, nine subjects reported a funny or great experience while five did not fully enjoy the experience. Regarding the quality of the display, two subjects mentioned good quality while three found the quality (or resolution) not as expected or not sufficient to provide a truly immersive experience. These findings explain the modest experience of enjoyment reported by our MOSs.

Subjects complained that it was tiring to have the two hands lifted towards their face in order to control the HMD. This complaint emphasizes the high likelihood of tiredness and lack of attention from subjects. We can address this issue by introducing a controller, such as a game-pad, in order to handle the interactions with the content (e.g. toggling and vote). Subjects also gave indications to improve the immersion and/or isolation experience: one mentioned than he was "experiencing content as a ghost instead of being physically there", another stated that "it would be more immersive to zoom in and out". Despite the soundness of these comments, several are outside of the scope of this study.

Some subjects indicated that the duration of the test was correctly set as they were starting to get bored and/or annoyed when visualizing the two or three last stimuli. Considering this comment, a test session (including explanations and training, if applicable) should not exceed 20 minutes as outlined in several similar subjective assessments. One subject mentioned his approach carried out during the experiment, especially regarding the visualization: "At the beginning, I tended to move in all directions and then, I progressively focused on areas of interest, typical of each content" and had the "tendency to navigate more horizontally than vertically, and then to balance in both directions". These are important indications for toggling information analysis.

A few subjects claimed that making a choice between the test and reference was a challenge. They found cases where they preferred the test content over the reference content in the bright regions while not so in the dark regions and vice-versa. This suggests that further work is needed in 360° TMO design to overcome such challenges.

5.3 Toggling analysis

The toggling feature is a key property of our methodology. This provides some inherent advantages over existing single stimulus evaluations including the ability to record the tracking and toggling information of the users. We believe it is possible to extract valuable information regarding regions of interests and subjects' behavior when comparing two omnidirectional HDR contents. This data could potentially be used to design 360°-TMOs and improve user experience in HMDs. We propose two methods to analyze this data:

- The first approach is based on the assumption that an effective comparison is operated exclusively where subjects have toggled. The toggling action undergone in a particular area most likely represents a region of interests which may influence the voting decision of the user. We refer to this method as toggling locations.

- The second approach takes into account the speed of the head movement of user which may give a better understanding of identified regions of interests. We assume that areas where head movement is slower or "fixated" is a region of interest and this is determined by the angular velocity from head tracking data. We refer to this method as fixation locations.

The following sections detail the toggling processing outlined above and the analysis of the results.

5.3.1 Toggling processing

The toggling processing of the first approach is rather simple and straightforward. It consists of displaying the toggling locations of every subjects for each content after having re-sampled those locations in cells of 1x1 degree.

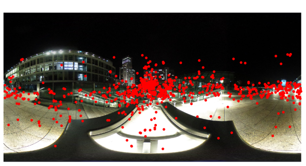
The implementation of the second approach is largely inspired from Upenik et al.[36] and proceeds as follows. The available raw data consists of head movement tracks and toggling information, in the form of two arrays of view-port center's yaw and pitch coordinates along with timestamps. The head movement tracks are first split into test and reference head movement tracks sets, according to which the test or reference stimuli was visualized during the track sample. We have defined fixation locations as being the locations where the angular velocity of an observer's head does not impact subjects ability to focus their attention on an object. The angular velocity is obtained by computing the first order time derivative of yaw and pitch coordinates. In order to remove any digital differentiation noise, the sets of head movement tracks were filtered, in advance of the derivation, with a second-order Butterworth low-pass filter with cutoff frequency of $f_c = 2$ Hz. The threshold defining if a track location is a fixation was set to 15 degrees per second, complying with the work of Upenik et al.,[36] and was verified as sound by an analysis on head tracking data angular velocity. The fixation locations of every subject were fused by summing all the locations in cells of 1x1 degree.

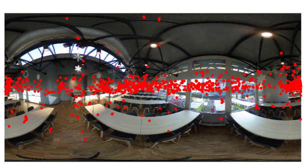
The processing performed in this study only considers the head movement positions. As a future extension of our work, more accurate fixations maps can be computed based on the prediction of the eye gaze fixation from head fixation locations, applying the work of Rai et al.[37] for instance. In addition, the selection of the threshold indicating if the angular velocity of observer's head prevents subjects focus of attention should be optimized.

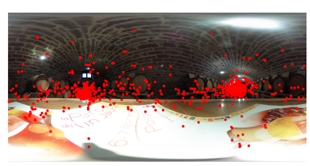
5.3.2 Toggling and Fixation locations analysis

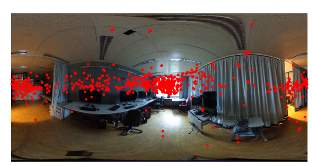
This section examines the results of the processing of toggling information previously-described. Figures 9 and 10 present the resulting toggling and fixation locations per content, respectively.

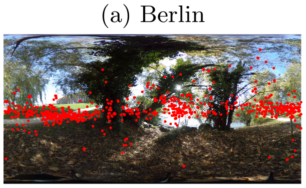
The resulting toggling and fixation locations per content and per operator are not included as the behavior of subjects across operators show very little variations, especially for toggling locations. This fact isn't in line with subjects behavior during 2D content visualization, investigated by Narwaria et al.[28] who indicate that every operator has a signature in visual attention maps. The toggling and fixation locations are thus not sufficient to investigate in depth subjects behavior, leading to the need of introducing eye gaze tracking or at least eye gaze prediction. This conclusion is also confirmed by the results of toggling and head fixation locations per content,

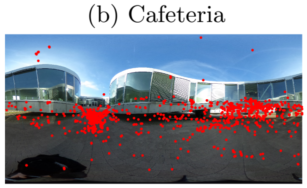

(a) Berlin

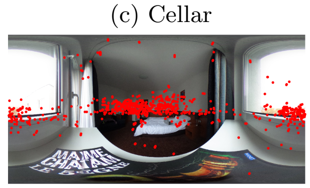

(b) Cafeteria


(c) Cellar


(d) Lab


(e) Lake


(f) Rolex


(g) Room

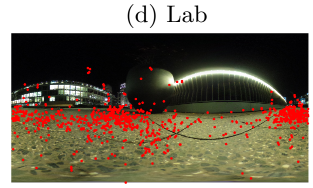
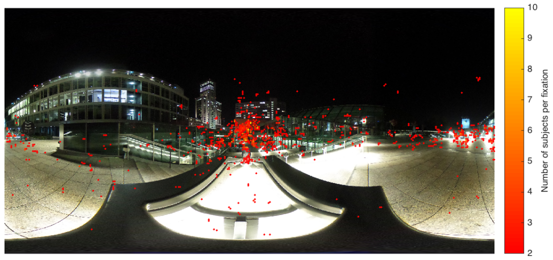

(h) Vase

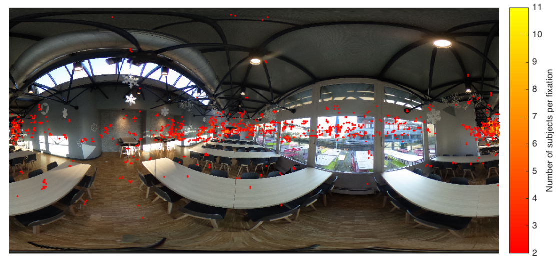
Figure 9: Toggling locations

per operator and per grade which demonstrates that toggling and fixation locations are not sufficient to indicate the reason behind subjects choices.

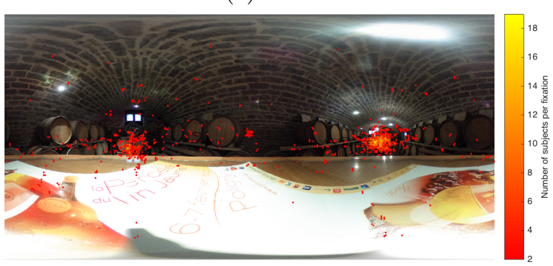
The insights provided in Figures 9 and 10 clearly show the equivalence of using toggling and fixation locations. However, no guidelines are drawn from this study as both methods have advantages and drawbacks. The toggling locations processing is not complex and emphasizes clear interest areas. However, the large variation in toggling coordinates prevents to see trends across subjects as well as makes the differentiation of two area of toggling more difficult in terms of importance. The fixation locations provide these insights as they indicate weighted fixation locations based on the number of subjects sharing a specific location. However, this method requires a more complex processing, which introduces a non-optimized parametrization (e.g. angular velocity threshold) and filtering. The results of this method can then be questioned as being less accurate or biased.



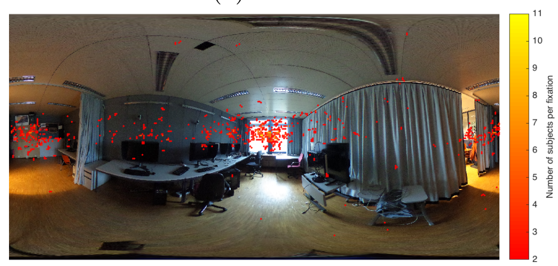
(a) Berlin



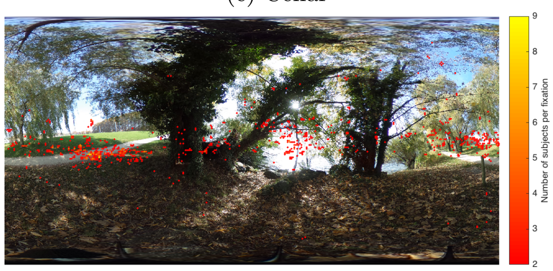
(b) Cafeteria



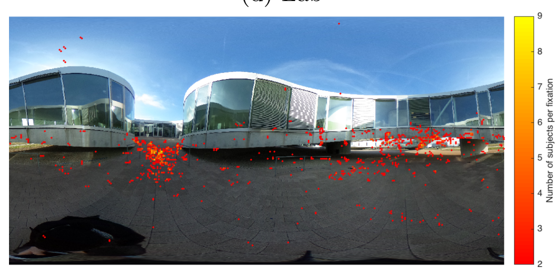
(c) Cellar



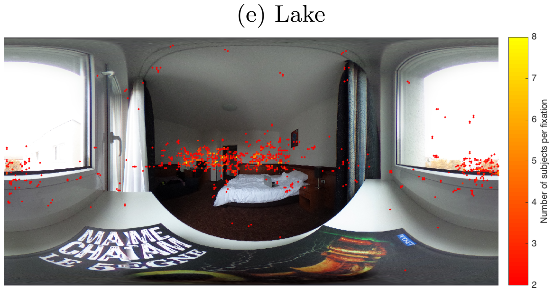
(d) Lab



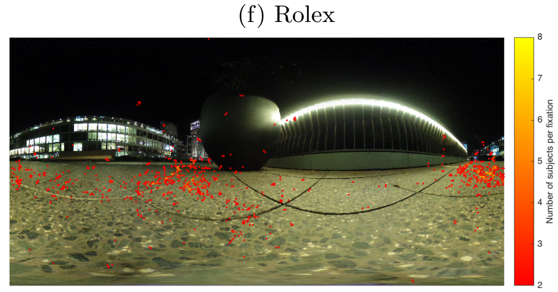
(e) Lake



(f) Rolex



(g) Room



(h) Vase

Figure 10: Fixation locations

In the remainder of this section, we are using both toggling and fixation locations results to draw conclusions about subject's behavior. It should also be mentioned that the term *areas of interest* refers, in the following, to the areas with a high density of toggling or fixation locations.

The first interesting result is that there is no influence of the initial view-port on the regions of interest. We observe this on the contents Cellar, Lake, Rolex and Vase where there is few or no toggling and fixation locations at the initial view-port location, indicating it is not an area of interest. In some cases, such as in Berlin, Cafeteria, Lab and Room images, the initial view-port is one of the areas of interests.

The locations emphasized in the results are mostly longitudinally distributed at a latitude close to the equator. Additionally, the main areas of toggling and fixation are also mainly centered at the equator latitude. In fact, the first quartile, average and third quartile of the set of pitch coordinates are 28.6, 1.9 and -23 respectively. This range of latitudinal head positions seems comfortable and natural for users.

While analyzing the areas of toggling and fixation, it seems that subjects are assessing their preference in terms of loss of information. The highlights of several contents, such as the outside information through the windows of the Lab and Room images as well as the building or the sky over the grass field in the Lake content, are clear toggling and fixation areas. These toggling and fixation locations correspond to areas over-exposed in SDR contents and which present more details in HDR contents. However, in some cases the over-exposed areas in SDR contents are not taken into consideration. For example the reflection of the light on the lake or windows in the Lake and Rolex contents respectively. Also, the objects that serve as light sources such as the light bulbs in the Cafeteria and Cellar images and lit rooms in the buildings seen on the in night contents were not considered by users. In such cases the loss of information is usually considered by expert viewers in order to have a faithful representation of reality. A possible explanation is that, for naive observers, the importance of the lost information of specular and light emitting areas is not significant compared to the other areas in the entire omnidirectional scene.

One can also notice the trend whereby observers focus more often on distant details than on close ones. In the Cellar content, the two clear areas of interest are the two extremities of the cellar. Overall, in all contents, the main areas of toggling and fixation coincide with the locations of the furthest objects. This subjects behavior is highly interesting and can lead to new visual attention models as well as provide clues for the development of 360°-dedicated TMOs. Thus, this finding needs to be confirmed by further analysis with more various contents, as well as with contents with a larger dynamic range and of course on another sample of population.

To summarize, the toggling and fixation locations information are equivalent and are not precise enough to identify the variations of subjects behavior depending on operator used. Also, the initial view-port has no influence on the subjects viewing experience. We also observed that subjects tend to browse through the content longitudinally with a reduced latitudinal range. Finally, the assessments of subjects seem to be based on the loss of information in specific areas and indicate subjects tendency to focus on distant objects.

6. CONCLUSION

In this paper, we propose to assess the importance of developing HDR for omnidirectional contents as well as proposing the directions of improvement for dedicated TMOs. An evaluation of an end-to-end HDR pipeline using a consumer device has been carried out. We introduce a publicly available dataset composed of 43 multi-exposure images, acquired with a consumer camera, also including the HDR reconstructions and the SDR versions resulting from four well-known off-the-shelf TMOs and one exposure fusion algorithm. Eight of those contents, carefully selected, were used to conduct evaluation tests. A new subjective test methodology, enabling pairwise comparisons for omnidirectional content is introduced. The toggling permits to switch between the assessed stimuli and the reference. Our results show that none of the evaluated operators show a clear increase of perceived quality, however the exposure fusion, as well as the linear TMO to a lesser extent, show promising results. In addition, discrepancies across operators and contents lead to the identification of a need for dedicated TMOs for omnidirectional content. As future work, evaluation of TMOs with professional contents can be considered. The analysis performed on our self-made questionnaire emphasized that *Details in bright areas*, *Overall contrast* and the *Naturalness of the scene* are important criteria to consider during the assessment. Some subjects reported slight sickness effects mostly felt as *Nausea* and *General discomfort*. This fact confirms the proper design of our methodology. Furthermore, we analyzed the toggling information to draw important conclusions regarding human perception while comparing tone mapped omnidirectional images on a head mounted display. This work aims to lay a basis for the future development of HDR imaging for omnidirectional representations.

ACKNOWLEDGMENTS

Touradj Ebrahimi and Anne-Flore Perrin acknowledge funding from Swiss State Secretariat for Education, Research and Innovation (SERI) for participation to the Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 643072 - Network QoE-Net. The work of Cambodge Bist has been supported by the Institute of Research and Technology b<>com, dedicated to digital technologies and funded through the French National Research Agency (ANR) Investment referenced ANR-A0-AIRT-07.

REFERENCES

[1] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K., [*High dynamic range imaging: acquisition, display, and image-based lighting*], Morgan Kaufmann (2010).

[2] Mann, S. and Picard, R., [*Being undigital with digital cameras*], MIT Media Lab Perceptual (1994).

[3] Debevec, P. and Malik, J., "Recovering high dynamic range radiance maps from photographs," in [*Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97)*], 369–378, ACM Press/Addison-Wesley Publishing Co. (1998).

[4] Mertens, T., Kautz, J., and Van Reeth, F., "Exposure fusion: A simple and practical alternative to high dynamic range photography," in [*Computer Graphics Forum*], **28**(1), 161–171, Wiley Online Library (2009).

[5] Debevec, P., "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in [*ACM SIGGRAPH 2008 classes*], 32, ACM (2008).

[6] Banterle, F., Artusi, A., Debattista, K., and Chalmers, A., [*Advanced High Dynamic Range Imaging: Theory and Practice*], AK Peters (CRC Press), Natick, MA, USA (2011).

[7] Ostiak, P., "Implementation of hdr panorama stitching algorithm," in [*Proceedings of the 10th CESCG Conference*], 24–26 (2006).

[8] Mantiuk, R., Krawczyk, G., Mantiuk, R., and Seidel, H.-P., "High dynamic range imaging pipeline: Perception-motivated representation of visual content," in [*Human Vision and Electronic Imaging XII*], Rogowitz, B. E., Pappas, T. N., and Daly, S. J., eds., *Proceedings of SPIE* **6492**, SPIE, San Jose, USA (February 2007).

[9] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J., "Photographic tone reproduction for digital images," *ACM transactions on graphics (TOG)* **21**(3), 267–276 (2002).

[10] Mantiuk, R., Daly, S., and Kerofsky, L., "Display adaptive tone mapping," in [*ACM Transactions on Graphics (TOG)*], **27**(3), 68, ACM (2008).

[11] Mantiuk, R., Myszkowski, K., and Seidel, H.-P., "A perceptual framework for contrast processing of high dynamic range images," *ACM Transactions on Applied Perception (TAP)* **3**(3), 286–308 (2006).

[12] Hausner, A. and Stamminger, M., "Tone Mapping in VR Environments," in [*Eurographics 2004 - Short Presentations*], Alexa, M. and Galin, E., eds., Eurographics Association (2004).

[13] Yang, X. S., Zhang, L., Wong, T.-T., and Heng, P.-A., "Binocular tone mapping.," *ACM Trans. Graph.* **31**(4), 93–1 (2012).

[14] Mikamo, M., Mori, K., Raytchev, B., Tamaki, T., and Kaneda, K., "Binocular tone reproduction display for an hdr panorama image," in [*Proceedings of the ACM Symposium on Applied Perception*], 131–131, ACM (2016).

[15] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H., "Evaluation of tone mapping operators using a high dynamic range display," in [*ACM Transactions on Graphics (TOG)*], **24**(3), 640–648, ACM (2005).

[16] Yoshida, A., Blanz, V., Myszkowski, K., and Seidel, H.-P., "Perceptual evaluation of tone mapping operators with real-world scenes," in [*Electronic Imaging 2005*], 192–203, International Society for Optics and Photonics (2005).

[17] Kuang, J., Yamaguchi, H., Johnson, G. M., and Fairchild, M. D., "Testing hdr image rendering algorithms," in [*Color and Imaging Conference*], **2004**(1), 315–320, Society for Imaging Science and Technology (2004).

[18] Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J., "Evaluation of tone mapping operators for hdr-video," in [*Computer Graphics Forum*], **32**(7), 275–284, Wiley Online Library (2013).

[19] Melo, M., Bessa, M., Debattista, K., and Chalmers, A., "Evaluation of tone-mapping operators for hdr video under different ambient luminance levels," in [*Computer Graphics Forum*], **34**(8), 38–49 (2015).

[20] Narwaria, M., Da Silva, M. P., Le Callet, P., and Pepion, R., "Single exposure vs tone mapped high dynamic range images: a study based on quality of experience," in [*Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*], 2140–2144, IEEE (2014).

[21] Yu, M., Lakshman, H., and Girod, B., "A framework to evaluate omnidirectional video coding schemes," in [*Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*], 31–36, IEEE (2015).

[22] Zakharchenko, V., Choi, K. P., and Park, J. H., "Quality metric for spherical panoramic video," in [*SPIE Optical Engineering+ Applications*], 99700C–99700C, International Society for Optics and Photonics (2016).

[23] Upenik, E., Rerabek, M., and Ebrahimi, T., "A testbed for subjective evaluation of omnidirectional visual content," in [*32nd Picture Coding Symposium*], (EPFL-CONF-221560) (2016).

[24] ITU-R BT.2035-0-201307, [*A reference viewing environment for evaluation of HDTV program material or completed programmes*] (July 2013).

[25] Akyüz, A. O. and Gençtav, A., "A reality check for radiometric camera response recovery algorithms," *Computers & Graphics* **37**(7), 935–943 (2013).

[26] Boitard, R., Bouatouch, K., Cozot, R., Thoreau, D., and Gruson, A., "Temporal coherency for video tone mapping," in [*SPIE Optical Engineering+ Applications*], 84990D–84990D, International Society for Optics and Photonics (2012).

[27] ITU-R BT.500-13, [*Methodology for the subjective assessment of the quality of television pictures.*] (January 2012).

[28] Narwaria, M., Da Silva, M. P., Le Callet, P., and Pépion, R., "Effect of tone mapping operators on visual attention deployment," in [*SPIE Optical Engineering+ Applications*], 84990G–84990G, International Society for Optics and Photonics (2012).

[29] Mansilla, W. A. C., *Quality of Aesthetic Experience and Implicit Modulating Factors*, PhD thesis (2013).

[30] Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G., "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology* **3**(3), 203–220 (1993).

[31] Bradley, M. M. and Lang, P. J., "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry* **25**(1), 49–59 (1994).

[32] Reinhard, E., Kunkel, T., Marion, Y., Brouillat, J., Cozot, R., and Bouatouch, K., "Image display algorithms for high-and low-dynamic-range display devices," *Journal of the Society for Information Display* **15**(12), 997–1014 (2007).

[33] Eilertsen, G., Mantiuk, R., and Unger, J., "Real-time noise-aware tone mapping," in [*ACM Transactions on Graphics*], **34**(6), 198–1, Association for Computing Machinery (ACM) (2015).

[34] de Borda, J. C., "Mémoire sur les élections au scrutin," (1781).

[35] Cobb, S. V., Nichols, S., Ramsey, A., and Wilson, J. R., "Virtual reality-induced symptoms and effects (vrise)," *Presence: teleoperators and virtual environments* **8**(2), 169–186 (1999).

[36] Upenik, E. and Ebrahimi, T., "A simple method to obtain visual attention data in head mounted virtual reality," in [*IEEE International Conference on Multimedia and Expo 2017*], (EPFL-CONF-227457) (2017).

[37] Rai, Y., Le Callet, P., and Guillotel, P., "Which saliency weighting for omnidirectional image quality assessment?," in [*9th International Conference on Quality of Multimedia Experience (QoMEX 2017)*], IEEE (2017).